

## Aberystwyth University

### *VORFFIP-Driven Dock*

Segura, Joan; Marín-López, Manuel Alejandro; Jones, Pamela F.; Oliva, Baldo; Fernandez-fuentes, Narcis; Keskin, Ozlem

*Published in:*  
PLoS One

*DOI:*  
[10.1371/journal.pone.0118107](https://doi.org/10.1371/journal.pone.0118107)

*Publication date:*  
2015

*Citation for published version (APA):*

Segura, J., Marín-López, M. A., Jones, P. F., Oliva, B., Fernandez-fuentes, N., & Keskin, O. (Ed.) (2015). VORFFIP-Driven Dock: V-D<sup>2</sup>O<sup>2</sup>CK, a Fast and Accurate Protein Docking Strategy. *PLoS One*, 10(3), [e0118107]. <https://doi.org/10.1371/journal.pone.0118107>

#### **Document License** CC BY

#### **General rights**

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400  
email: [is@aber.ac.uk](mailto:is@aber.ac.uk)

RESEARCH ARTICLE

# VORFFIP-Driven Dock: V-D<sup>2</sup>OCK, a Fast and Accurate Protein Docking Strategy

Joan Segura<sup>1‡a</sup>, Manuel Alejandro Marín-López<sup>2</sup>, Pamela F. Jones<sup>1</sup>, Baldo Oliva<sup>2</sup>, Narcis Fernandez-Fuentes<sup>1‡b\*</sup>

**1** Leeds Institute of Molecular Medicine, School of Medicine, University of Leeds, Leeds, LS9 7TF, United Kingdom, **2** Structural Bioinformatics Lab (GRIB-IMIM), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, 08003 Barcelona, Catalonia, Spain

‡a Current address: Department of Macromolecules, Spanish National Center for Biotechnology, CNB-CSIC, Darwin 3, Madrid, 28049 Spain

‡b Current address: Institute of Biological, Environmental and Rural Science, Aberystwyth University, Aberystwyth, SY23 3EB, United Kingdom

\* [narcis.fernandez@gmail.com](mailto:narcis.fernandez@gmail.com)



## OPEN ACCESS

**Citation:** Segura J, Marín-López MA, Jones PF, Oliva B, Fernandez-Fuentes N (2015) VORFFIP-Driven Dock: V-D<sup>2</sup>OCK, a Fast and Accurate Protein Docking Strategy. PLoS ONE 10(3): e0118107. doi:10.1371/journal.pone.0118107

**Academic Editor:** Ozlem Keskin, Koç University, TURKEY

**Received:** September 16, 2014

**Accepted:** December 27, 2014

**Published:** March 12, 2015

**Copyright:** © 2015 Segura et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Data for this study are available on the Harvard Dataverse Network (DOI: 10.7910/DVN/28610), <https://thedata.harvard.edu/dvn/dv/VD2OCK-B04>.

**Funding:** This work was supported by Research Councils UK (RCUK) under the RCUK Academic Fellowship program (NFF) and a PhD scholarship awarded by the University of Leeds (JS). BO acknowledges support from the Spanish Ministry of Economy and Competitiveness; grant number BIO2011-22568 and MAML a PhD scholarship awarded by the Generalitat de Catalonia (FI-DGR2012). The funders had no role in study design,

## Abstract

The experimental determination of the structure of protein complexes cannot keep pace with the generation of interactomic data, hence resulting in an ever-expanding gap. As the structural details of protein complexes are central to a full understanding of the function and dynamics of the cell machinery, alternative strategies are needed to circumvent the bottleneck in structure determination. Computational protein docking is a valid and valuable approach to model the structure of protein complexes. In this work, we describe a novel computational strategy to predict the structure of protein complexes based on data-driven docking: VORFFIP-driven dock (V-D<sup>2</sup>OCK). This new approach makes use of our newly described method to predict functional sites in protein structures, VORFFIP, to define the region to be sampled during docking and structural clustering to reduce the number of models to be examined by users. V-D<sup>2</sup>OCK has been benchmarked using a validated and diverse set of protein complexes and compared to a state-of-art docking method. The speed and accuracy compared to contemporary tools justifies the potential use of VD<sup>2</sup>OCK for high-throughput, genome-wide, protein docking. Finally, we have developed a web interface that allows users to browser and visualize V-D<sup>2</sup>OCK predictions from the convenience of their web-browsers.

## Introduction

One of the most prevalent challenges in the post-genomic era is the charting and description of the protein networks that underpin cellular functions. Large-scale interactomic experiments (e.g. [1,2]) sought to describe the protein interactions that occur in cells, and albeit valuable, most of the information derived from these experiments does not provide the underlying structural, atomic details of the interactions. These details are central in order to realize the full potential of interactomic data in rational approaches such as the development of novel drugs to target protein

data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** Narcis Fernandez-Fuentes is a PLOS One Editorial Member Board and that this does not alter the authors' adherence to PLOS ONE Editorial policies and criteria.

interfaces[3] or understanding the effect of mutations[4], for example. Computational methods can be used to derive structural models of protein complexes (reviewed in [5] and references therein), which can then be used as the starting point for further research approaches.

Protein docking represents one such computational approach. Protein docking is an active field of research; shown by the number of participants in the regular Critical Assessment of PRediction of Interactions (CAPRI) exercises[6] and the number of publications devoted to the field. Protein docking methods can be broadly divided in two groups: unbiased (or *ab initio*) and biased (or data-driven) approaches and implementations of both have been described in the scientific literature (e.g. [7–18]). The major difference between *ab initio* and data-driven docking is that the latter group restricts the sampling of docking to selected region(s) of the proteins, whereas in the former group the sampling of the docking space is not restricted. The constraints to guide data-driven docking can be derived from either experimental methods (e.g. Hydrogen-Deuterium exchange data[19]) or computational predictions (e.g. binding site predictions [8]).

In this work, we present the development of a high-throughput computational docking strategy: V-D<sup>2</sup>OCK, which combines protein-binding site prediction and data-driven docking. V-D<sup>2</sup>OCK also includes a clustering step to reduce the number of docking poses while preserving the conformational richness of the sampling. Our results show that V-D<sup>2</sup>OCK is a competitive and faster approach than *ab initio* docking and successfully samples the docking space generating near-native docking poses. The clustering step resulted in only limited decrease in performance while substantially reducing the number of docking solutions, a desirable characteristic in a day-to-day use of this technology. V-D<sup>2</sup>OCK is accessible as a web application at <http://www.bioinsilico.org/VD2OCK>. The web-server includes a bespoke and interactive graphic viewer that allows users to examine and manipulate the docking poses using the web-browser.

## Material and Methods

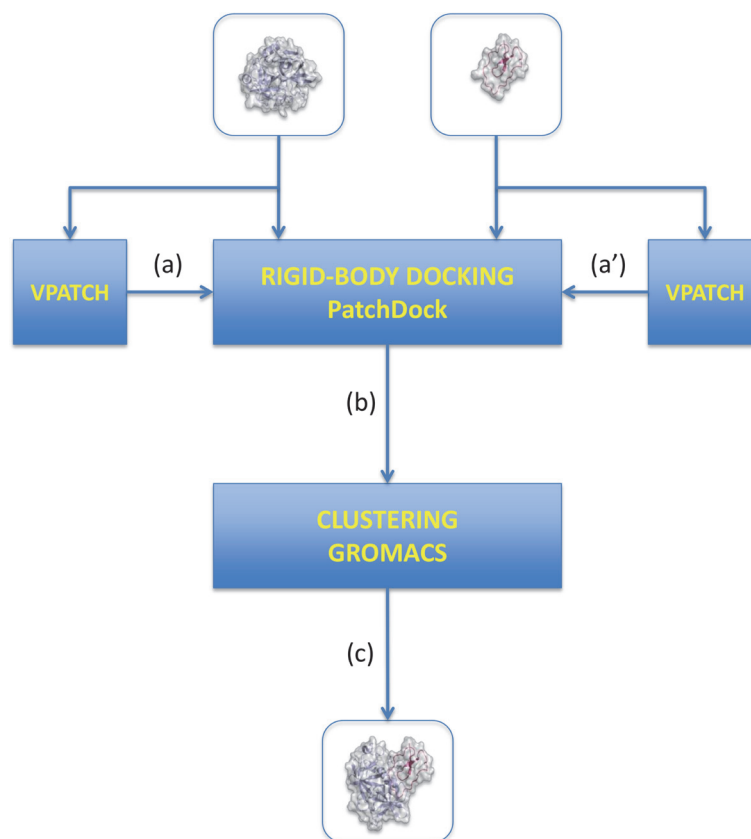
### Datasets

The benchmarking of V-D<sup>2</sup>OCK was performed using Benchmark v4.0 [20] referred here as the B04 set. B04 was specifically compiled to test docking methods and it consists of 176 complexes classified in: rigid-body, medium difficulty and difficult cases depending on the structural changes upon complex formation. The atomic structures for the proteins are available in both bound and unbound conformations.

In the case of the V-PATCH algorithm, a dataset referred to here as SOB4 was derived from an original set of protein complexes described in Ofra et al.[21] after removing any protein complexes whose SCOP superfamily [22] was represented in B04. This set was used to train VORFFIP [23], hence avoiding any bias between the training and testing set. The protein interfaces of the native complexes in B04 were determined using DIMPLLOT [24] on the bound complexes. The binding site prediction scores were computed using VORFFIP on the unbound structures.

### V-D<sup>2</sup>OCK algorithm

The V-D<sup>2</sup>OCK algorithm is composed of different steps that include the prediction of binding sites in proteins, sampling of the docking space using data-driven docking and the clustering of the docking poses to reduce the number examined (Fig. 1).



**Fig 1. VD<sup>2</sup>OCK workflow.** (a) (a') V-PATCH algorithm is used to define the protein binding sites based; (b) rigid-body docking is driven by interface predictions; and (c) clustering stage where dockings poses are structurally clustered and clusters' centroids selected as representatives.

doi:10.1371/journal.pone.0118107.g001

## From single residues to interaction patches: V-PATCH

The first step of the algorithm involves the delineation of the binding sites in both partners. VORFFIP [23] was used to assign scores to individual scores using the unbound structure that were then used to compute the interface patches by V-PATCH. VORFFIP scores are fed into V-PATCH to define explicit binding sites in a protein structure by an automatic and iterative clustering of residues that include: (i) the initial patch generation; (ii) patch selection; and (iii) patch extension.

In the initial patch generation, a new score named extended score:  $s_i^*$  (1) is calculated for each residue that includes VORFFIP's original score and the contribution of the environment scores as defined in our earlier work [23]. Let  $\{(a_k, s_k); k = 1, \dots, N\}$  be the residues and predicted scores of a given protein and  $\{a_j; j = 1, \dots, n\}$  neighbours of residue  $a_i$ , then  $s_i^*$  is defined as

$$s_i^* = 0.5[s_i' + \sum_{j=1}^n c_{ij}s_j'] \quad (1)$$

where  $c_{ij}$  is the contact strength between  $a_i$  and  $a_j$  and  $s'_j$  (2) is the normalized score calculated as

$$s'_j = \frac{s_j - m}{M - m} \quad (2)$$

being  $m = \min\{s_i; i = 1, \dots, N\}$  and  $M = \max\{s_i; i = 1, \dots, N\}$ .

The initial patches are then started with the residues with the highest scores and extended to any neighbouring residues until  $s'_i$  falls below a threshold  $\alpha$  or hard-average cut off. The parameter  $\alpha$  was calculated using the average of the extended scores for interface residues in the complexes of SOB4 dataset.

During the patch selection stage, redundant patches are removed. The list of patches is sorted by size, and any smaller patches that are also associated with a larger patch are removed, retaining only the largest patch. The last stage of the algorithm extends the patches to maximize the size of the interface patch by including neighbouring residues that were not selected in the previous rounds and whose extended score is above a certain threshold  $\beta$ . The parameter  $\beta$ , named the soft average cut-off, is calculated by computing the average of extended scores in the case of residues that are not part of protein interfaces in SOB4 dataset. An explicit pseudo code implementation of the algorithm is available in the supplementary material (S1 File).

## Data-driven docking and clustering of the docking space

The patches computed by V-PATCH are then used to guide the docking of protein partners. V-D<sup>2</sup>OCK utilizes PatchDock[25] [15] to perform the docking of the proteins. The list of residues conforming the patches identified by V-PATCH is given as an input to PatchDock.

The third stage of the algorithm is the structural clustering of the docking poses to reduce the redundancy and size. This method used the g\_cluster application, part of the GROMACS package [26]. The g\_cluster is executed with default parameters except for the RMSD cut-off, which is set up to 5 Angstroms (Ang), based on the threshold used in the CAPRI competition [6] to define a docking solution as medium accuracy. Thus, this ensures that all members within a cluster will have a similar RMSD if compared to the centroid.

## Scoring of docking models

Three different scoring functions were used to rank the docking models: (i) PatchDock native score[15]; (ii) the ES3DC potential, a distance and environment dependent knowledge-based statistical potential[27]; and (iii) ZRANK[28]. The complete set of docking complexes derived for the entire B04 using V-D<sup>2</sup>OCK is available as a compressed file (bzip2) at [http://www.bioinsilico.org/VD2OCK/PD\\_B4\\_results.tar.bz2](http://www.bioinsilico.org/VD2OCK/PD_B4_results.tar.bz2), upon request to the authors or at the Harvard Dataverse Network (S2 File).

## Statistical measures

Four widely used statistical measures were used to assess the performance: Recall (1), Precision (2), the Matthews Correlation Coefficient (MCC)(3), and the F1 score(4). Formally,

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FP)(TN + FN)}} \quad (3)$$

$$F1 = \frac{2TP}{2TP + FN + FP} \quad (4)$$

where TP is the number of true positives, TN true negatives, FP false positives, and FN the number of false negatives.

## Results and Discussion

### V-Patch algorithm

The V-PATCH algorithm was designed to define interface patches based on VORFFIP predictions[23]. The patches defined by V-PATCH were compared to the native interfaces of the protein complexes in the dataset B04. In 28 cases out 175 complexes, the predicted interface matched 80% or more of the native interface residues, in 82 cases the success ranged between 20 and 80% overlap and in 12 cases the overlap between the native and predicted interface was less than 5%. Having a high overlap between the predicted and native interface is highly desirable in data-driven docking, albeit not vital, since including only a few native contacts (i.e. low overlap) is usually enough and limitations can be corrected during the docking process as recently discussed [8]. On the other hand, over-predicting also presents disadvantages: it increases the search space and hence computational time and the number of docking solutions to rank is higher.

To fully assess the advantages of V-PATCH algorithm, the accuracy of predicted interface patches defined on the basis of a fixed threshold (both raw and normalized VORFFIP scores) and V-PATCH were compared. As shown in Table 1, V-PATCH performs better than fixed threshold for all the statistical measures: recall, precision, F1 scores and MCC. V-PATCH has the clear advantage that no thresholds need to be defined. Moreover, V-PATCH has been designed such that multiple, independent, patches on the surface can be defined, i.e. it can generate different, independent, binding sites.

**Table 1. Statistical performance of V-PATCH and fixed thresholds.**

Method	R(%)	P(%)	F1	MCC
V-PATCH	61	27	0.37	0.34
Fixed Threshold (raw)	60	22	0.32	0.29
Fixed Threshold (norm)	60	24	0.34	0.30

VPATCH and fixed thresholds using raw and normalized score were used to compare performance. Results are shown for (R) recall, (P) precision, the (F1) F1 score and (MCC) Matthews correlation coefficient

doi:10.1371/journal.pone.0118107.t001

## Sampling of docking space on a validated set: B04

The first question to address was the completeness of sampling of the docking space by V-D<sup>2</sup>OCK in order to understand whether near-native structural poses were generated. The performance of sampling was assessed in terms of the ligand-Root Mean Square Deviation, l-RMSD, adapting the scoring scheme from CAPRI[6]: high accuracy (three-stars), medium accuracy (two-stars), acceptable (one-star) or wrong. With an average number of around 4000 docking poses per complex, V-D<sup>2</sup>OCK yielded acceptable and medium quality structural models for over 70% of the cases; one case ranked as high-quality (the Falcipain-2 and Cystatin complex [PDB code 1yvb]) and in 30% of the cases the docking failed to sample any suitable conformation (see Table 2). Specific information on each individual protein including the theoretical minimum RMSD, i.e. best docking pose, is shown in S1 Table (supplementary material).

V-D<sup>2</sup>OCK generates an average of 1353 docking poses per interface and 4509 docking poses per complex. Given the large number of docking poses and the challenges it might present for routine use and downstream processing such as energy minimization, a clustering step based on structural similarity was devised. Different clustering cut-offs were explored to assess the impact on the quality of the sampling. The l-RMSD of the best docking poses was computed when: (i) considering all poses (no clustering), (ii) considering the centroids of all clusters; (iii) considering the centroids of the top 1000 clusters; (iv) considering the centroids of the top 200 clusters; (v) considering the centroids of the top 100 clusters; and (vi) considering the centroids of top 50 clusters as per PatchDock scoring function[25]. As shown in Table 2, increasing in clustering stringency results in a decrease in the quality of the models. This is due to the intrinsic structural variability among the models that belong to the same cluster as only the centroid is considered for calculation purposes. However, there is a clear advantage in the clustering as the number of poses reduces dramatically, thus reducing the number of models to be inspected, while the reduction in the quality of the models is lesser to some extent in comparison (e.g. no clustering vs. all poses). Moreover, all members of the cluster can be easily retrieved upon inspection of the structure of the centroid (see V-D<sup>2</sup>OCK web server).

Due to its nature, data-driven docking is less comprehensive than *ab initio* docking, i.e. data-driven docking directs the docking of receptor and ligand and thus restricts the search space. To further clarify the effect of the constraints imposed by the selection of interfaces and quality of the docking poses, we studied the relationship between the best l-RMSD and the

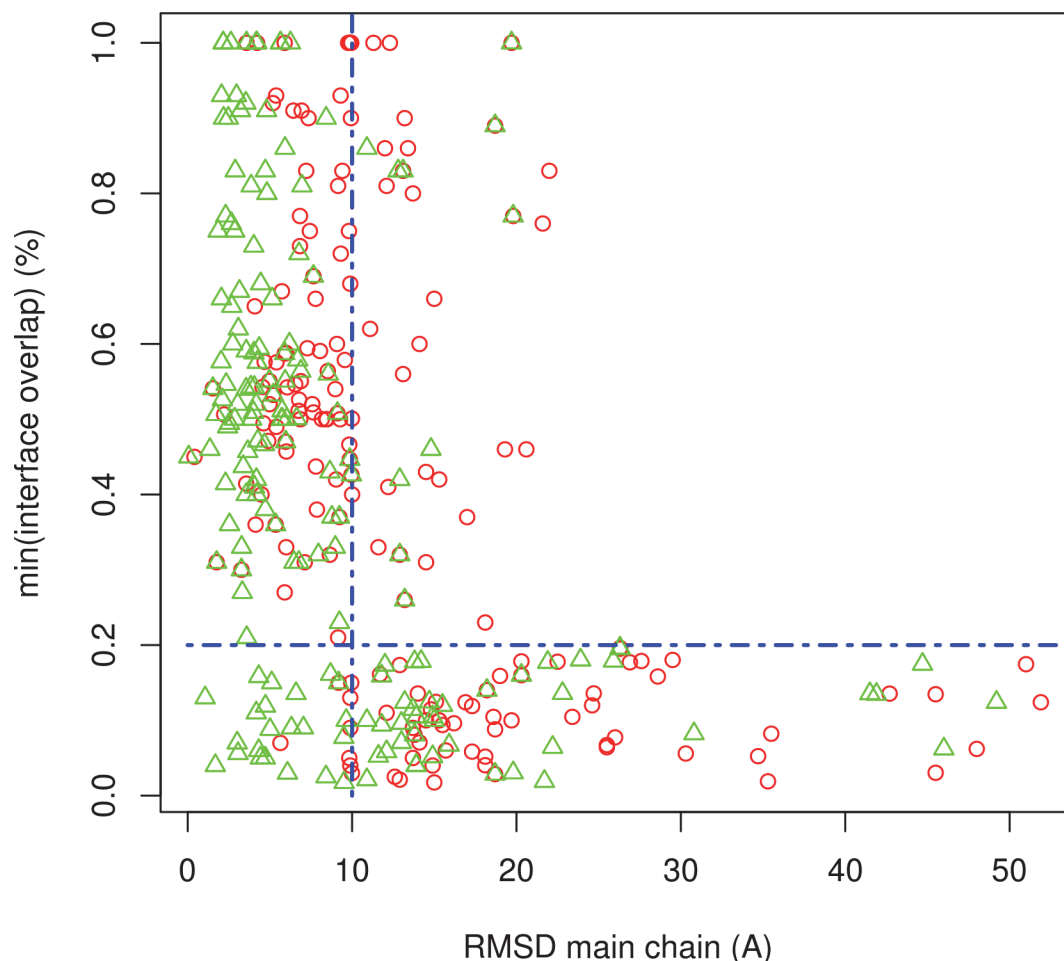
**Table 2. Effect of clustering in the quality of the models.**

# of solutions to cluster(a)	CAPRI evaluation system(b)				# Docking poses(c)
	A	B	C	D	
No clustering	1	46	75	53	4509
All clusters	0	21	93	61	898
1000	0	20	83	72	635
200	0	17	73	85	218
100	0	13	60	102	162
50	0	9	47	119	96

(a) Sets of solutions used: all poses, centroids for all clusters, centroids for the top 1000, centroids for the top 200, centroids for the top 100 and centroids for the top 50 clusters respectively. (b) CAPRI evaluation system where A, B, C and D represent the number of predictions considered as high-quality quality (three stars), medium quality (two stars), acceptable and wrong respectively are shown alongside the average number of docking models (c)

doi:10.1371/journal.pone.0118107.t002





**Fig 2. Relationship between l-RMSD (Å) and interface coverage (%).** RMSD was calculated using the main chain atoms. The interface coverage represents the lowest coverage of the predicted binding sites in either ligand / or receptor. Red empty circles and green empty triangles represent the best l-RMSD using all docking poses or the best poses among the top 200 clusters respectively.

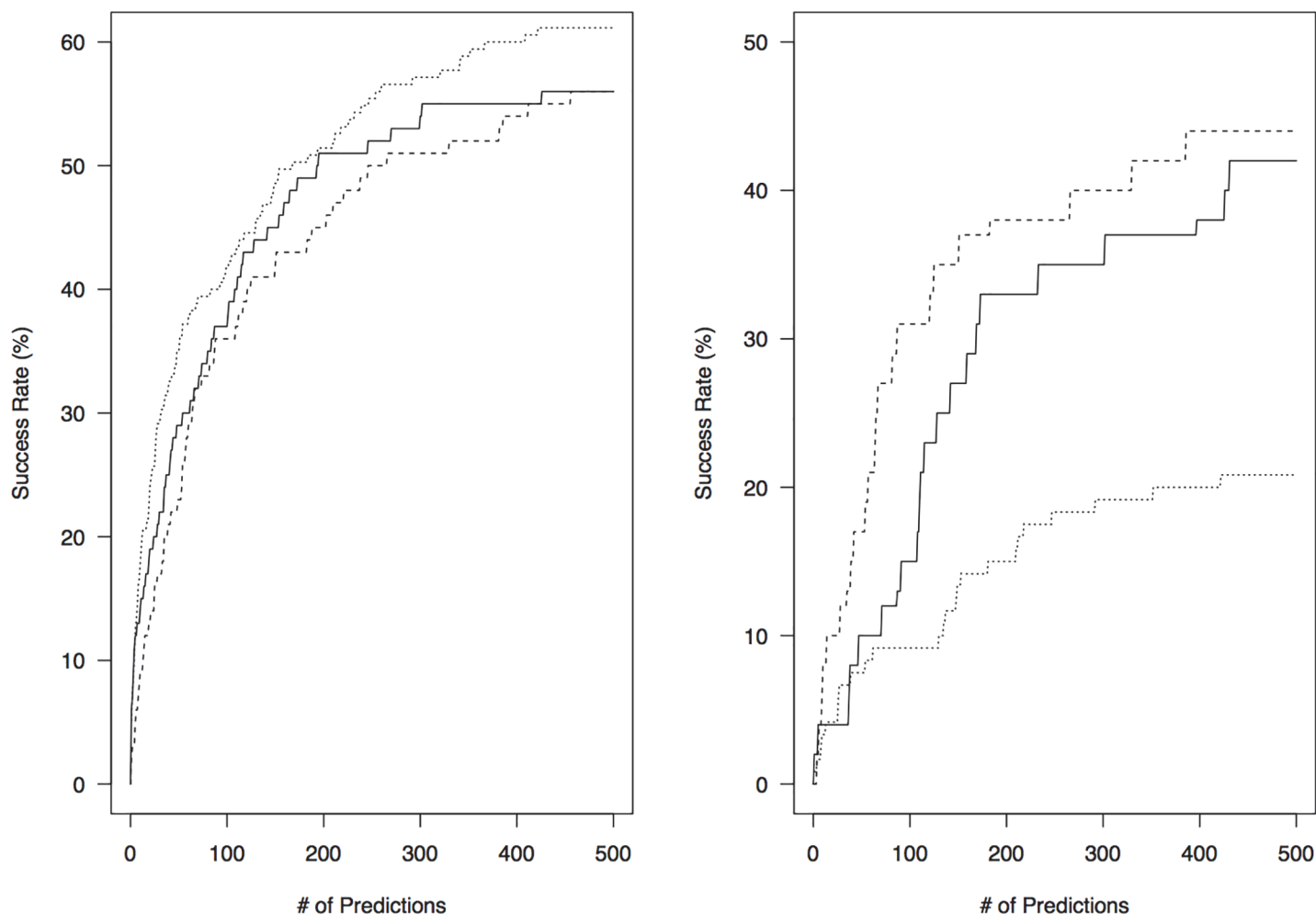
doi:10.1371/journal.pone.0118107.g002

overlap of the predicted and real interfaces (Fig. 2). As shown, only when the overlap of the predicted and native interface drops below 20% does the quality of docking models deteriorate substantially. Above 20% of interface overlap, V-D<sup>2</sup>OCK consistently samples docking poses below 10 Å l-RMSD. These results agree with those previously reported by de Vries et al, which show that inclusion of a low proportion of native contacts is usually sufficient as the docking process can correct for the actual orientation of the proteins [8].

### Comparing VD<sup>2</sup>OCK and a competitive *ab initio* docking algorithm: ZDOCK

From the previous analysis, it can be concluded that the sampling of the docking space is efficient and generates docking poses close to the native ones, even though the search is directed by the predicted interfaces. The performance of the method was then assessed in terms of successful predictions among the top N, N being the number of predictions being considered. Three different scoring functions were considered: PatchDock native score [15], the ES3DC potential [27], and ZRANK score [28].





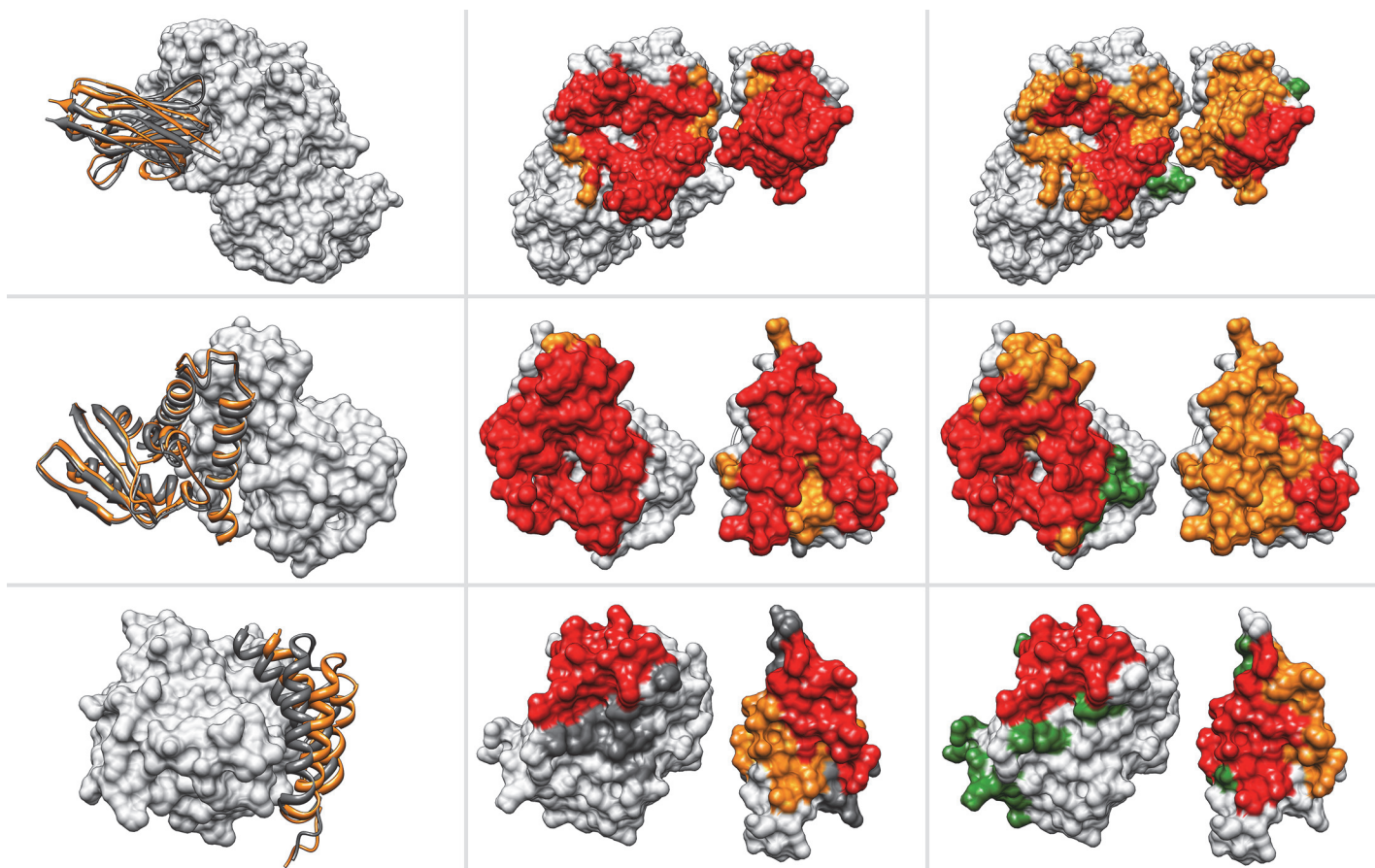
**Fig 3. Success rates for all test cases (left) and medium/difficult cases (right) on Benchmark v4.0.** PatchDock[15], ES3DC potential[27] and ZRANK scores[28] are shown as solid, dashed and dotted lines respectively.

doi:10.1371/journal.pone.0118107.g003

The success rate is around 55% when considering the top 500 poses (Fig. 3) and 69% when considering only rigid-body, or easy, cases (S1 Fig.). The scoring function that performed the best was ZRANK, followed by PatchDock and ES3DC potential. However, ES3DC, a coarse-grained statistical potential, outperformed both ZRANK and PatchDock native scores for flexible/difficult cases with a success rate close to 45% in comparison to 19% for ZRANK (Fig. 3). In general, the performance achieved is similar, if not higher, than ZDOCK[13], an *ab initio* protein docking method that was also benchmarked using the same dataset: BO4 [20] (for exhaustive comparison of success rate curves see Fig. 1 in [13].)

### Examples of predicted complexes using VD<sup>2</sup>OCK

Fig. 4 illustrates three different examples of predicted complexes one for each of the classes defined in BO4 [20], i.e. easy, medium and difficult. These classes are defined depending on the level of conformational change upon formation of the protein complex: easy class is similar to rigid body docking (i.e. no conformation change); medium and difficult class implies conformational changes in the monomers upon binding. The first example, member of the ‘easy’ class in BO4, is the protein complex formed by a camelid VHH domain bound to the porcine



**Fig 4. Examples of structural models.** Rows (from top to bottom) show the comparison between native and predicted structures of protein complexes: camelid VHH domain and porcine pancreatic alpha-amylase (PDB code 1kxq)[29], BET3 and TPC6B core of TRAPP (PDB code 2cfh)[30], and Pol II epsilon and Hot proofreading complex (PDB code 2ido)[31]. Columns (from left to right) show: 1) the structure of native and predicted complex where receptor is depicted in surface (grey) and receptor as ribbon representation (native: dark grey; predicted: orange). 2) Surface representation of both receptor (left) and ligand (right) and the overlap (red) between native (dark grey) and predicted (orange). 3) Surface representation as in 2) showing the overlap (red) between predicted interface (green) and docking interface (orange).

doi:10.1371/journal.pone.0118107.g004

pancreatic alpha-amylase[29]. The second example represents a case of medium difficulty as per B04 classification and corresponds to the protein complex formed by a human Bet3 and Tpc6B of the transport protein particle complex[30]. Finally, the third example, the epsilon subunit of *E.coli* polymerase III in complex with the Hot protein[31], corresponds to the 'difficult' class.

For all the cases described above VD<sup>2</sup>Ock derived docking poses that closely resembles the structure of the respective native complexes. The superimposition of the native and predicted complexes (first column) and the overlap between the native and predicted interfaces (second column) show that the predicted structures closely resembles the native complex. In addition, the third column shows that the overlap between the V-PATCH predicted residues and the predicted complex interfaces is not total, showing that the docking is in fact correcting the final interface. These observations agree with previous[8] and our own observations in this work that shows that even with a low overlap between predicted and native interfaces the docking process can correct for the missing information(Fig. 2 and S1 Table).

## V-D<sup>2</sup>OCK web server

Although the computing time is not a measure of quality, it can limit the applicability of the method. In its current implementation, V-D<sup>2</sup>OCK requires 36 CPU hours on a standard desktop to complete the entire B04 dataset, i.e. 12 minutes per complex on average, including the prediction of interfaces, docking and clustering. Given the speed of the algorithm, V-D<sup>2</sup>OCK has been implemented as a web-application and predictions can be derived in real time (<http://www.bioinsilico.org/VD2OCK>). The web server provides a user-friendly interface to execute the docking algorithm and to analyse and visualize the structural models. As described, the number of potential docking poses can be large, even after applying the clustering step. The web application, however, features a bespoke viewer that allows easy navigation and visualization among the structural models. The structural models can be also sorted according to different criteria, which include the PatchDock (default), ZRANK and ES3DC scores, contact surface area, and cluster size. Finally, the coordinates of the docking poses, both centroids and poses within clusters, can be also downloaded.

## Conclusions

Here we describe V-D<sup>2</sup>OCK, a data-driven docking strategy that integrates V-PATCH, PatchDock[15] and a final clustering step. As shown, the method is able to sample suitable docking conformations even with low coverage of the native interfaces. The clustering step greatly reduces the number of docking poses with a limited impact on the quality of the models, facilitating the analysis and visualization of the docking solutions. We have explored different scoring functions and depending on the nature of the conformational change upon formation of the protein complex, ES3DC coarse-grained statistical potential performs better than ZRANK energy-based function. Finally, V-D<sup>2</sup>OCK is accessible via a web application, which features a bespoke molecular visualizer that allows users to easily and conveniently analyse, visualize and download the structural models of protein complexes. Moreover, users can select additional scoring functions and/or download the models generated by V-D<sup>2</sup>OCK and apply the scoring function of choice.

## Supporting Information

**S1 Fig. Success rates for easy cases (rigid-body) on Benchmark v4.0.** PatchDock[15], ES3DC potential[27] and ZRANK[28] scores are shown as solid, dashed and dotted lines respectively.  
(TIFF)

**S1 File. Pseudo-code implementation of VPATCH algorithm.**  
(DOCX)

**S2 File. URLs to download docking decoys derived for B04 using VD<sup>2</sup>OCK.**  
(DOCX)

**S1 Table. VD<sup>2</sup>OCK predictions for protein complex on Benchmark v4.0.** Columns represent the PDB code (first column), overlap between predicted and native interface in receptor (%; second column), overlap between predicted and native interface in ligand (%; third column), and l-RMSD (Ang) of the best docking pose (fourth column). The rest of the columns are grouped in 5 blocks of 3, each showing the l-RMSD (Ang) for the top scoring pose using PatchDock (PD)[15], ES3DC[27] and ZRANK (ZR)[28] scores within the TOP 1, TOP 10, TOP 50, TOP 100 and TOP 200 respectively. Blue, yellow and green blocks of the table show the easy, medium and difficult cases according to Benchmark v4.0 classification[20].  
(DOCX)

## Acknowledgments

We thank ZRANK, PatchDock and GROMACS authors for making their software freely available to the scientific community. NFF thanks Dr Gendra for insightful comments to the manuscript.

## Author Contributions

Conceived and designed the experiments: JS BO NFF. Performed the experiments: JS MAML. Analyzed the data: JS MAML BO NFF PFJ. Contributed reagents/materials/analysis tools: MAML BO. Wrote the paper: JS MAML BO NFF PFJ.

## References

1. Ewing RM, Chu P, Elisma F, Li H, Taylor P, et al. (2007) Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* 3: 89. PMID: [17353931](#)
2. Gavin A, Aloy P, Grandi P, Krause R, Boesche M, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. *Nature* 440: 631–636. PMID: [16429126](#)
3. Mullard A (2012) Protein-protein interaction inhibitors get into the groove. *Nature reviews Drug discovery* 11: 173–175. doi: [10.1038/nrd3680](#) PMID: [22378255](#)
4. Wang X, Wei X, Thijssen B, Das J, Lipkin SM, et al. (2012) Three-dimensional reconstruction of protein networks provides insight into human genetic disease. *Nature biotechnology* 30: 159–164. doi: [10.1038/nbt.2106](#) PMID: [22252508](#)
5. Stein A, Mosca R, Aloy P (2011) Three-dimensional modeling of protein interactions and complexes is going [ ]omics. *Current Opinion in Structural Biology* 21: 200–208%U Available: <http://www.sciencedirect.com/science/article/pii/S0959440X11000078>. doi: [10.1016/j.sbi.2011.01.005](#) PMID: [21320770](#)
6. Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, et al. (2003) CAPRI: a Critical Assessment of PRe-dicted Interactions. *Proteins* 52: 2–9. PMID: [12784359](#)
7. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Research* 32: W96–99. PMID: [15215358](#)
8. de Vries SJ, Bonvin AM (2011) CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PloS one* 6: e17695. doi: [10.1371/journal.pone.0017695](#) PMID: [21464987](#)
9. Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125: 1731–1737. PMID: [12580598](#)
10. Gabb HA, Jackson RM, Sternberg MJ (1997) Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J Mol Biol* 272: 106.
11. Lesk VI, Sternberg MJ (2008) 3D-Garden: a system for modelling protein-protein complexes based on conformational refinement of ensembles generated with the marching cubes algorithm. *Bioinformatics* 24: 1137–1144. doi: [10.1093/bioinformatics/btn093](#) PMID: [18326508](#)
12. Lyskov S, Gray JJ (2008) The RosettaDock server for local protein-protein docking. *Nucleic Acids Research* 36: W233–238. doi: [10.1093/nar/gkn216](#) PMID: [18442991](#)
13. Pierce BG, Hourai Y, Weng ZP (2011) Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. *PloS one* 6.
14. Ritchie DW, Venkatraman V (2010) Ultra-fast FFT protein docking on graphics processors. *Bioinformatics* 26: 2398–2405. doi: [10.1093/bioinformatics/btq444](#) PMID: [20685958](#)
15. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ (2005) PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Research* 33: W363–367. PMID: [15980490](#)
16. Torchala M, Moal IH, Chaleil RA, Fernandez-Recio J, Bates PA (2013) SwarmDock: a server for flexible protein-protein docking. *Bioinformatics* 29: 807–809. doi: [10.1093/bioinformatics/btt038](#) PMID: [23343604](#)
17. Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research* 34: W310–314. PMID: [16845016](#)
18. Li B, Kihara D (2012) Protein docking prediction using predicted protein-protein interface. *BMC Bioinformatics* 13: 7. doi: [10.1186/1471-2105-13-7](#) PMID: [22233443](#)

19. Xiao H, Verdier-Pinard P, Fernandez-Fuentes N, Burd B, Angeletti R, et al. (2006) Insights into the mechanism of microtubule stabilization by Taxol. *Proceedings of the National Academy of Sciences of the United States of America* 103: 10166–10173. PMID: [16801540](#)
20. Hwang H, Vreven T, Janin J, Weng Z (2010) Protein-protein docking benchmark version 4.0. *Proteins* 78: 3111–3114. doi: [10.1002/prot.22830](#) PMID: [20806234](#)
21. Ofra Y, Rost B (2003) Analysing six types of protein-protein interfaces. *J Mol Biol* 325: 377–387. PMID: [12488102](#)
22. Hubbard TJ, Murzin AG, Brenner SE, Chothia C (1997) SCOP: a structural classification of proteins database. *Nucleic Acids Res* 25: 236. PMID: [9016544](#)
23. Segura J, Jones PF, Fernandez-Fuentes N (2011) Improving the prediction of protein binding sites by combining heterogeneous data and Voronoi diagrams. *BMC bioinformatics* 12: 352. doi: [10.1186/1471-2105-12-352](#) PMID: [21861881](#)
24. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* 8: 127. PMID: [7630882](#)
25. Duhovny D, Nussinov R, Wolfson H (2002) Efficient Unbound Docking of Rigid Molecules. In: Guigó R, Gusfield D, editors. *Algorithms in Bioinformatics*: Springer Berlin Heidelberg. pp. 185–200.
26. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26: 1701–1718. PMID: [16211538](#)
27. Feliu E, Aloy P, Oliva B (2011) On the analysis of protein-protein interactions via knowledge-based potentials for the prediction of protein-protein docking. *Protein Sci* 20: 529–541. doi: [10.1002/pro.585](#) PMID: [21432933](#)
28. Pierce B, Weng Z (2007) ZRANK: reranking protein docking predictions with an optimized energy function. *Proteins* 67: 1078–1086. PMID: [17373710](#)
29. Desmyter A, Spinelli S, Payan F, Lauwereys M, Wyns L, et al. (2002) Three camelid VHH domains in complex with porcine pancreatic alpha-amylase. Inhibition and versatility of binding topology. *J Biol Chem* 277: 23645–23650. PMID: [11960990](#)
30. Kummel D, Muller JJ, Roske Y, Henke N, Heinemann U (2006) Structure of the Bet3-Tpc6B core of TRAPP: two Tpc6 paralogs form trimeric complexes with Bet3 and Mum2. *J Mol Biol* 361: 22–32. PMID: [16828797](#)
31. Kirby TW, Harvey S, DeRose EF, Chalov S, Chikova AK, et al. (2006) Structure of the Escherichia coli DNA polymerase III epsilon-HOT proofreading complex. *J Biol Chem* 281: 38466–38471. PMID: [16973612](#)